所谓影响关系有待商榷：对温忠麟等人（2024）的评论*

葛枭语

(北京大学心理与认知科学学院暨行为与心理健康北京市重点实验室, 北京 100871)

摘　要　温忠麟等人（2024）在《心理学报》发文聚焦长期以来用法模糊的“影响”一词，并提出“影响关系”概念，或将改变汉语心理学界接下来的语用实践。然而，该文存在若干疑点：“影响”一词无论汉英、无论大众或学界，均视为因果语言，难以被作者说服相信不传达因果意涵；“影响关系”缺失明确定义，似与因果关系并无本质差异；混淆目标与实现目标的手段，由于在手段上无法为因果目标提供因果证据，便创设一个不同于因果和相关关系的第三种目标，是奇怪的。更重要的是，共变且有方向性的变量关系并非如作者所言未被命名，一直被学界称为“预测”。据此，此评论旨在提醒学界同仁考虑描述相关情形的其它可能办法、审慎决策是否将所谓“影响关系”引入研究。

关键词　影响关系，相关关系，因果关系，影响因素，预测性研究

分类号　B84-0

温忠麟等人（2024）在《心理学报》发文聚焦心理学界长期以来用法模糊的“影响”一词，建设性地指出这种模棱两可的现状应有所改变，并创造性地提出了“影响关系”这一新颖概念，富于启发。然则此事关系重大，或将改变汉语心理学界接下来的语用实践，不得不三思后行。详阅该文，不免发现若干疑点，遂作此评论，愿与原作者和学界同仁共同探讨。

1　该文存在的疑点

温忠麟等人（2024）指出，区分影响关系、因果关系、相关关系很有必要，将影响关系误解为后二者会带来严重后果，但此概念尚未有明确界定。为此，他们主张，相关关系是指满足了“因和果共同变化（共变）”之准则的变量关系，影响关系是指在此基础上进一步满足了“因在果之前发生（方向性）”之准则的变量关系，因果关系则是指在前两者基础上进一步满足了“排除因果联系外的其它解释（排他解释）”之准则的变量关系。他们举例，研

收稿日期：2024-10-15
通信作者：葛枭语, E-mail: gexyu@foxmail.com

究者在横断研究中让一些学生报告一下自己的情绪和学习成绩（GPA），只要回归系数显著，他们主张研究者无须排除任何替代解释就可以将这个结果描述为“情绪影响 GPA”。接下来，我将详细阐述我的四点疑问：从写一篇概念文章的必备要件来看，该文缺失对“影响关系”的明确定义；从概念本身来看，该文提出的推论准则并不能有效区分影响关系与因果关系；从概念向读者传递的意涵来看，“影响”一词无论汉英、无论大众或学界，均视为因果语言，难以被作者说服相信不传达因果意涵；从本质问题来看（最重要的一点），该文混淆了目标与实现目标的手段。

1.1 所谓影响关系究竟是一种怎样的关系

最显而易见的疑点是，“影响关系”缺失明确定义。既然作者旨在主张影响关系是不同于相关关系、因果关系的新概念，按理说应当至少明确旧概念是指什么、新概念是指什么，以便读者比对三者异同。然而，通观全文，作者只提推论准则（即共变性+方向性），却从不给相关关系、影响关系、因果关系下定义。这个问题肯定不是作者忘了，因为《心理学报》公开的审稿意见显示，审稿人 1 意见 2、审稿人 2 意见 1 均指出概念界定问题，但作者仍以推论准则（有方向性的相关关系）来回答，未下定义，尽管在摘要中将推论准则称为定义。

为什么推论准则不可视同定义？一个很现成的案例是因果关系。既然温忠麟等人（2024）在摘要中“将影响关系定义为有方向性的相关关系”，那么因果关系亦可仿此方法被定义为有方向性的而且排除了因果联系外的其它解释的相关关系。问题来了，需要排除什么“其它解释”才能推论因果呢？X 和 Y 之间的关系存在多种解释可能性，其中一种符合因果关系之定义、其它均不符合，因此，凡是使得 X 和 Y 之关系不符合因果关系定义的解释，均须排除。那么，因果关系的定义是什么呢？这样看来，因果关系的“定义”就是排除了一切使其不能成为因果关系之解释的关系。这种尴尬的循环，提示我们不可将推论准则视同定义。唯有定义明晰，读者才能判断，作者所提议的推论准则能不能恰如其分地起到推论影响关系的作用；否则，读罢全文，读者仍不知影响关系究竟是在说什么。此外，定义虽可存在争议，但应有较大共识；推论准则虽可以有常用且公认的，但仍在很大程度上是开放的、且是从属于定义的，例如，作者指出的推论因果关系的三个准则就并非唯一判别标准、尽管确实常用（例如 Pirlott 和 MacKinnon（2016）就枚举了三准则以外的许多因果推论标准）。

让我们来看看 Bailey 等人（2024）的定义：因果效应（causal effect）是指，对于一个群体、一个群体的一个部分或一个人而言，在干预 X 的不同条件下，变量 Y 的假设差异。其中，干预（intervention）是指，对一个模型中的一个或多个变量所做出的任何有意改变（经常是出于揭示因果效应的目标）。Rubin（2005）也给出了类似的定义。简而言之，若改变

X，研究者假设 *Y* 会有所不同，那么 *X* 与 *Y* 被构建为因果关系。再来看看温忠麟等人（2024）对影响关系的理解，例如，“但说到影响，通常的理解是一个事件对另一个事件的作用，是有方向性的”。既然一个事件都已经对另一个事件起了作用，所谓“起作用”必然是第二个事件发生了某些变化，那么请问，这种关系和因果关系有何不同？另一个能够窥见作者理解的例证是关于风险因素和保护性因素的讨论，作者指出后者“可以简单理解为可以降低负面结果发生的影响因素”。既然一个因素已经可以降低结果变量的发生概率，那么请问，这种关系和因果关系有何不同？作者未为影响关系提供明确定义，就我个人而言，不易想出明明 *X*“影响”着 *Y*、但 *X* 无论怎么改变都不会使 *Y* 产生丝毫差异的情形。除非作者能够给出与因果关系存在本质差异的定义，否则因果关系与影响关系之别似站不住脚。

1.2　所谓影响关系真的不同于因果关系吗

当然，尽管缺失定义，但温忠麟等人（2024）并非没有表达如何区分因果关系与影响关系，他们提出，二者根本区别在于要不要遵守第三条推论准则，该准则指出，应当排除因果联系外的其它解释；因此，若不排除，则属于影响关系。如前所述，实际上这条准则在很大程度上是一则“兜底条款”，可以装下非常多更加细化的要求（凡是使得 *X* 和 *Y* 之关系不符合因果关系定义的解释，均须排除），且因研究问题而异，当然，其中最核心的就是，要防止反向因果或第三变量的存在导致了“非因果性”的共变（Pirlott & MacKinnon, 2016）。并不是说不允许有反向因果或第三变量，例如有些 *X* 和 *Y* 的关系已被证实是互为因果（例如 Tam & Inzlicht, 2024），但研究者担心的是所观察到的共变完全是由于这些混淆。

举个例子，我想知道双十一大促开始前的预热期广告投入金额是否影响双十一期间的销售金额，这是一个营销从业者的常见疑问。显然，二者在时间上是有方向性的（无论测量意义上、还是概念意义上）；若回归一下，也不难发现共变——广告投得多，销售额就多。然而，现在让我们想像一个极端的情形：某天某平台突然系统故障，导致只收取了广告费用、而没有任何广告真的被投放给了消费者。雪上加霜的是，甚至在这种极端情况下，共变仍然可以被观测到——大商家大品牌的营销预算多、销售目标也高，广告自然投得多；纵使没有广告，不出意外，它们也无疑会在大促期间斩获比小商家更大规模的销售成果。这时，数据研究员向老板汇报他们的分析结论：广告投入金额影响了销售金额。老板怒道：我刚听技术部门复盘了故障，根本就没有任何广告真的被投出去！这里分析出的“影响”是怎么个“影响”？幸运的是（不幸的是），大多数研究问题不会有一个技术部门跳出来警告研究者绝无可能存在影响，人类只能仰仗有限的研究手段来做推论。Bartram 等人（2024）举了一个有趣的例子：身高影响词汇量。为满足“方向性”要求，不妨将这个例子改造为：身高影响一

年后的词汇量。若按温忠麟等人（2024）的“定义”，要证实这种“影响”非常容易，只需要把刚出生的婴儿和青壮年放在同一个样本来做检验就可以了，既有共变、又有方向、无须排除替代解释，可下结论。默许研究者将这种结果描述为“身高影响一年后的词汇量”，令人不安。

这些例子想要传达的是，温忠麟等人（2024）把因果推论的三准则误解为互相独立，实际远非如此：若不排除替代解释（例如，可能是商家本身的生意规模同时影响了大促前的广告投入和大促中的销售额，可能是年龄同时影响了身高和词汇量），则第二条推论准则的所谓方向性也无从证得、“影响关系”无从谈起；若排除了替代解释，则影响关系与因果关系之异无从谈起，替代解释的排除手段固有优劣高下之分、但第三条准则本身难以担起区分影响关系与因果关系的重任。

1.3 读者能够相信“影响”不指称因果关系吗

因果关系与影响关系之界限暧昧，究其原因，可能是由于未准确把握“影响”一词的词义。《现代汉语词典》对动词“影响”的释义是：“对别人的思想或行动起作用。”温忠麟等人（2024）已在文中指明了“影响”一词对应的英语表达：impact、affect、influence。Merriam-Webster 对动词 impact 的主要释义是 to have a direct effect or impact on，对动词 affect 的主要释义是 to produce an effect upon (someone or something)和 to act on and cause a change in (someone or something)，对动词 influence 的主要释义是 to affect or alter by indirect or intangible means 和 to have an effect on the condition or development of。这些释义无不使用了非常明确的因果语言（cause and effect）。

非但普通大众，学术研究者也持有类似理解。Haber 等人（2022）要求 47 名流行病学、统计学、医学、经济学、心理学博士、博士后、教师对各类学术表达的因果意涵打分（无、弱、中、强）。对条形图的肉眼观察表明，cause 的因果意涵固然最强（中和强的比例约为 100%），但 affect、impact、influence 也毫不逊色（约为 90%）。荷兰期刊《社会指标研究》在其社论中明确将“influence”列入“只能传达因果意涵的语言（language that can only have a causal meaning）”，并警告投稿人在（简单的）横断研究中使用这种修辞策略将败坏手稿（Bartram et al., 2024）。遗憾的是，温忠麟等人（2024）并未向有可能采纳该文建议的研究者披露“影响”一词可能给他们的投稿和出版带来的这些不利后果。

由此观之，温忠麟等人（2024）试图说服普通读者和学界同仁相信“影响”一词不传达因果意涵，这是一个充满挑战性的目标，违背了“影响”一词的本来词义。纵使因果关系和相关关系外真的存在其它概念，影响关系也绝非这个新概念的恰当名称。

1.4 未能排除替代解释，到底是以独特的变量关系为目标，还是仅仅属于手段上的妥协

最重要的一点（这也是所有疑点的本质根源）：该文混淆了目标与实现目标的手段。就目标而言，区别无非是研究者是否旨在探究因果关系；瞄准因果关系的研究，在手段方面，为因果关系所提供的证据可能存在不同的强度（Hernán, 2018）。

温忠麟等人（2024）提出影响关系的初衷是，许多研究虽然在目标上想证明 *X* 影响 *Y*，但在手段上无法提供严格的因果性证据。为此，便创设一个不同于因果关系和相关关系的第三种目标，是奇怪的，因为这类研究本质上仍抱持着因果性目标。根据因果关系的定义，如果一项研究的理论目标是"若改变 *X*，研究者假设 *Y* 会有所不同"，则该研究瞄准了因果关系，无论研究者是否真的去改变了 *X*，也无论这个 *X* 在现实中是否有望被人为改变。据此，温忠麟等人（2024）为影响关系举的很多例子，就其目标而言，无非都是因果关系，例如教师教学水平影响学生成绩。研究者心中所想的是，若改变教师教学水平，学生成绩按理说会有所不同；若结果符合预期，还可能在讨论中建议通过提升教师教学水平来提高学生成绩，这是因果性的议题。研究者能否排除替代解释——这是温忠麟等人（2024）提议的影响关系与因果关系的唯一区别——仅仅是手段层面的问题，不会使研究者怀揣的目标发生变化：无论手段，他们就是想解释为什么学生成绩有高有低，原因之一可能是教师教学水平不同。

事实上，许多在手段上并非随机化实验的研究，其意欲解答的问题都是因果性的：例如，若大学少排早八点的课，能否提振学生 GPA（Yeo et al., 2023）；若在电影中埋下更多反转，能否取得更大成功（Knight et al., 2024）；若不感染新冠，能否避免神经精神症状风险的升高（Kim et al., 2024）；若提升某个家庭的社会经济地位，能否改善孙辈的人格特质（Martin & Donnellan, 2021）；若激活某地的社会流动性、减轻阶层固化，能否减少当地暴力犯罪（Mann et al., 2024）；若降低某地的新自由主义程度，能否看到当地政府和居民更加支持抗疫行为（Liu et al., 2024）。尽管这六个案例均未对 *X* 做实验操纵，其中一些 *X* 甚至在现实生活中都很难以人的意志为转移（更别提在研究中操纵了），但是，这些研究的理论建构和现实关切无疑是因果性的——或许，可以通过改变 *X*，来提升好的 *Y* 或避免坏的 *Y*——这也正是研究意义所在。由于研究者在具体手段上一时难以为因果目标提供充分证据，便辩解称他们本就不想解答因果性问题，是本末倒置；相反，大部分心理学实证研究都怀揣着因果性目标，意欲回答：为什么人们的心理和行为会有种种不同，是否改变某因素就能使人们的心理或行为向好变化。

尽管以因果为目标，但研究者常常面临现实的限制，不论是 *X* 真就无法被操纵、还是经费精力不足以支撑操纵性实验，不同研究在手段上给出的因果证据强度是不同的，而且很

难被类别化。只有在手段意义上，温忠麟等人（2024）被审稿人建议删去的 0~1 连续轴概念才是合适的：1 代表了实证证据使人们完全有信心相信因果关系，0 代表了完全没有信心（但似乎不能等价于相关关系）。许多努力可以帮助一项研究从 0 走向 1，例如控制变量、匹配、断点回归、工具变量、纵向追踪、交叉滞后、人内人间对比、直接检验替代解释、双重差分、孟德尔随机化分析、双胞胎或兄弟姐妹设计、自然主义实验、随机化实验等等，都可以令一项研究比单纯非操纵性横断调查的普通回归分析在因果推论方面更有说服力，因为它们多多少少帮读者排除了一些可能质疑因果关系的替代解释（例如反向因果、对比苹果和梨的谬误、第三变量等），尽管某些策略的单独使用仍可能偏 0（例如单纯使用控制变量很难说是提供了因果证据）。随机化实验并非一劳永逸地卡位于 1，实际上，不同组别的异质性、不同组别缺乏测量不变性、*X* 的理论概念与实际操纵手段之间的间隙、肥手干预、干扰、溢出、实验者效应等等因素都有可能伤害我们对因果关系的信心（Bailey et al., 2024）——即，让我们怀疑，若改变 *X*，*Y* 并不会像研究者所相信的那样发生变化。

厘清目标与手段，使温忠麟等人（2024）对于中介分析前提无法被满足的担忧也消弭了。他们认为，只要将某些原本被称为相关关系的研究重命名为影响关系，就能使这些相关性研究不会违反中介模型是因果模型的本质要求、可以名正言顺地实施中介分析。对此，不得不借用 Hernán（2018）的标题来回应，"科学委婉用语不能改善观测性数据的因果推论"。事实上，如果一个模型在目标（及其理论建构）上是因果的，即使在手段上一时无法提供因果证据，也不妨碍中介分析。当然，作者需要承认、读者需要注意，因果目标可能在很大程度上未被手段所满足，存在许多替代解释未能排除，例如可能是因变量影响中介变量。相反，如果一个模型在理论上就说不通因果关系，那么，纵使换十个名号也于事无补，不可能因此就满足了中介模型是因果模型的本质要求。中介分析说到底是要排除三种可能性：*X* 变化不会引起中介变量的变化，中介变量的变化不会引起 *Y* 的变化，中介变量对 *Y* 的影响独立于 *X* 对 *Y* 的影响（Ge, 2023）。试问，如果 *X* 虽然号称"影响"*Y*，但由于不是因果关系，*X* 无论如何变化都不会使 *Y* 发生丝毫变化（请注意是理论层面上、不是统计层面上），那么二者"绝缘"，中介有何立足之地？

2 另一种可能的描述方式

那么，既然共变的且在时间上有方向性的关系不宜被叫作影响关系，是否有名称可以命名呢？尽管温忠麟等人（2024）在文中反复申述影响关系这一称谓的必要性、强调这种关系未被命名，但事实上，这种关系一直有明确且公认的叫法——"预测"。

2.1　目标层面

Hamaker 等人（2020）将科学研究分为三类（当然，非其独创，类似的分类方法在很多地方都可找到）：描述性研究；预测性研究，是指根据在时间上靠前的一个或多个变量来预测在时间上靠后的特定变量的值；解释性研究，主要目标是了解潜在的因果机制并制定干预措施。举几个例子就可以快速了解为什么有的研究者只关心预测目标、不关心因果目标。假如我要设计一种商业医疗保险，我所能知道的信息就是每一位客户的体检报告，我想知道的是，这位客户在保险生效以后有多大的可能性罹患合约范围内的疾病，这样我才能通过保费保额的设计确保我不会亏损。作为保险设计者，我的目标显然不是，通过改变客户的身体健康（*X*）来降低他们找我报销的概率（*Y*），我只想知道这个概率是多少。又如，公共卫生研究经常把大量人口统计学指标放入模型，计算在这些指标上处于某个水平的群体是否比其他群体更易罹患某种疾病。如果研究者发现，某种性别、某种性取向、某种社会经济地位的群体，抑郁症风险更高，他们显然不可能建议这些群体干脆改一个别的性别、性取向或地位，而是在资源有限的情况下优先对他们实施宣导教育或干预措施——这个干预措施并不是改变 *X*（性别、性取向、地位）。（注：不是说但凡以这些变量为自变量就一定是预测性研究，例如，一项研究关心的是扶贫——类似于改变社会经济地位——能否减轻心理疾病风险，纵使没有真去扶贫、仅测量，这也是因果性研究，需要关注的是研究者实际关心的问题、而非这里作为示例的变量。）预测性研究最时髦的实例大约是人工智能（不论预测式的、还是生成式的），尽管 AI 训练模型远比心理学常用的线性和逻辑等预测模型要复杂得多，但在目标方面并无二致：作为一位 AI，我要在训练样本里面充分学习概率关系，我从不纠结于任何一对 *X* 和 *Y* 有无因果关系，我努力的方向只是，下次有人输入一些 *X*，即便这些 *X* 在我的训练样本中无法找到直接对应的案例，我也可以猜到输出哪种 *Y* 最符合输入者的要求。

在这些例子中，研究者对因果目标（“若改变 *X*，研究者假设 *Y* 会有所不同”）毫无兴趣，他们不想改变 *X*，只想知道：如果我已知 *X*，那么 *Y* 在多大程度上可以被准确预测。因此，以预测为目标的研究者不太需要担心替代解释的混淆（Bailey et al., 2024），他们只关心增减哪些预测源、使用哪种量化手段可以提升预测精度（Martin & Kushwaha, 2024），以及如何选取样本能助力他们把从训练样本中捕捉到的关系有效推广至更宽泛的人群和环境（或边界条件在哪里）。例如，研究者利用银行交易记录（所反映出的人类行为）来预测一个人的人格特质（Gladstone et al., 2019），无须关心是花销影响人格、还是人格影响花销、又或是第三变量同时影响二者，只要已知花销、能在可接受的精确度上猜中人格，预测性目标就达成了。通俗点说，他们不担心太阳从东边升起并非放诸寰宇而皆准，只要不离开地球，

就能猜中明天太阳从哪儿升起；也就是说，只要训练样本与应用场景之间的人群和环境特征没有根本性区别，他们积累的知识就能够使他们做出对 *Y* 的有效预测。现在让我们回想一下前文的案例，如果数据研究员的分析目标改变成，用广告投入金额来预测销售金额，我们很容易相信，纵使无因果，这个预测性结论也可以在未来几年得到有效复现。又如，你过年回家遇到一位完全不认识的亲戚，需要通过猜测对方的词汇量来决定自己如何跟其说话，你想起自己曾做过一项研究，在 0~30 岁人群的样本中，身高可以预测词汇量，尽管不存在因果，但这项研究此时格外有用：身高不足 1 米，童言童语最合适；身高跟你差不多，你用平时的词汇交谈即可，不担心对方听不懂。但是，如果你不巧遇到了一位 30~60 岁的人，你想把从 0~30 岁样本中得到的预测性结论（身高越高，词汇量越大）推广到这个人身上，预测有效性就变得很可疑了。有心理学研究者甚至呼吁，应当多关注预测性研究，而把解释作为次要目标（Yarkoni & Westfall, 2017）；当然，如何抉择还是应取决于研究者感兴趣的议题。

2.2 手段层面

如果一个研究者已经选定因果（而非预测）作为目标，但其理论论证和实证证据仅能表明 *X* 先于 *Y* 且二者共变，无法实证性地排除威胁因果关系的大部分替代解释，那么，与其称呼为“*X* 可以影响 *Y*”，然后教育大众这里的影响不是你们日常理解的那种影响，何不直接称呼为“*X* 可以预测 *Y*”呢？实际上大部分研究者也正是这样做的（一些来自顶级心理学期刊的例子包括程亚华等，2023；Engstrom et al., 2024；Lu et al., 2021；Meyer et al., 2023；Rotella et al., 2021；Shen & Shoda, 2021；Stasi et al., 2024；Wright & Jackson, 2023）。在这里使用“预测”，并不意味着这项研究就是以预测为目标了，而仅仅是对所做的分析及其结果做出了事实描述。在词义上，《现代汉语词典》将预测解为“预先推测或测定”，Merriam-Webster 将 predict 解为 to declare or indicate in advance，无须重新定义，大众读者就可领会研究者想传达的意思。

2.3 目标与手段

在这里，我并非在相关关系之外提出了新概念，相反，无论是瞄准了预测性的目标、还是表明了预测性的结果，无疑都是非因果性的。我把上面提到的种种区分，以图形的方式在图 1 表示。在目标层面，首先的分野还是在于研究者想不想探究因果关系（“若改变 *X*，研究者假设 *Y* 会有所不同”）。若不想，则该研究不需要按照因果关系的标准来考核，进一步的区分可以按照方向性：若研究者关心方向，则为预测性研究；若不关心方向，则属于纯粹的关联性研究（简单称为相关关系似乎有些歧义）。若研究者在目标上致力于探究因果关系，纵使在手段上无法充分证明（纵使并未真的人为改变 *X*），也应按照因果关系的标准来

考核研究设计的质量。若能在手段上尽可能多地排除威胁因果关系的替代解释，则读者对因果关系的信心更强烈（靠近图 1 数轴的 1 端）；若不能，则更薄弱（靠近图 1 数轴的 0 端），那么，就不应简单宣称因果结论（尽管可以谈论因果意涵），而应当用不特定于因果关系的语言来描述结果和结论：例如，*X* 与 *Y* 有关（be associated with、association/relationship between/of ... and/with ...）；越……越……；*X* 较高的人，*Y* 更高；某组比另一组的 *Y* 更高；*X* 可以预测 *Y*（predict）等。

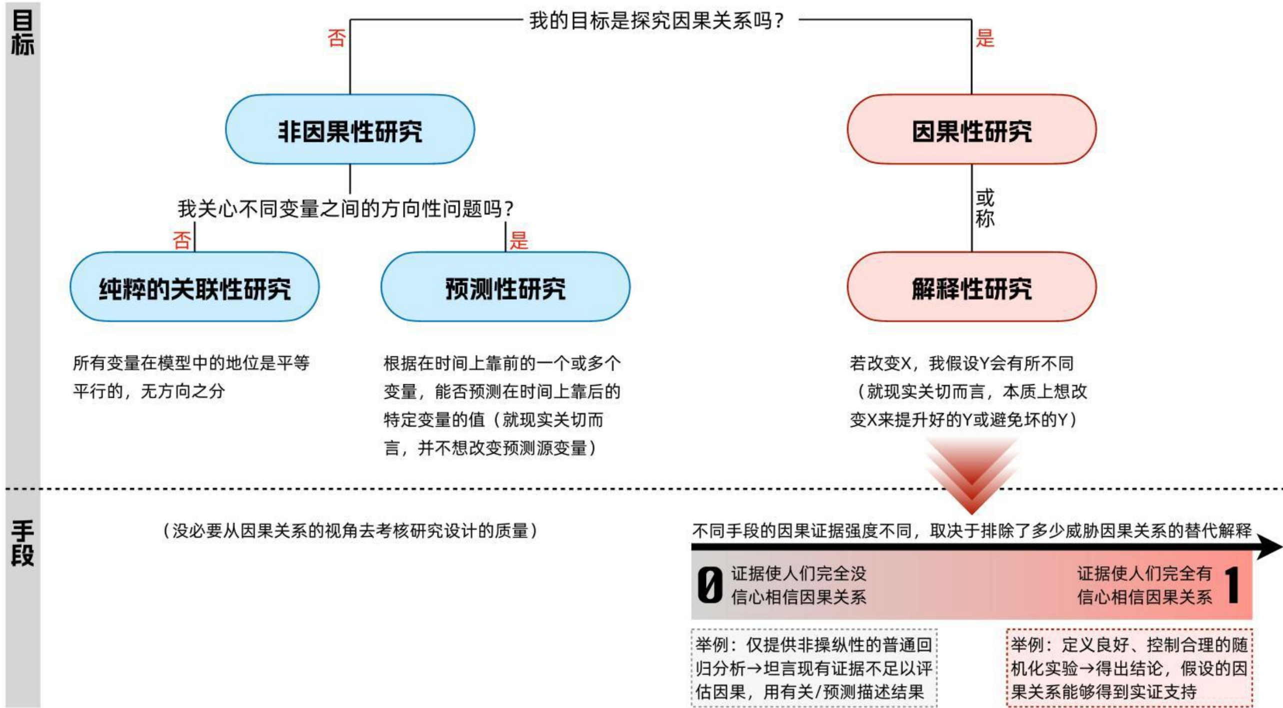


图 1　研究目标与手段

3　可考虑的行动建议

尽管影响关系之论存在诸多疑点，但是，温忠麟等人（2024）确实准确地切中了当前研究实践中的真实问题——“影响”一词在使用上存在诸多混淆之处，而有方向性、但未能提供充分的因果证据的研究在报告结果时遇到困境。针对这些问题，如下建议（以及随附的相关文献）可供作者、审稿人、读者参考。

其一，明确研究目标是什么。（A）纯粹的关联性研究：常见案例是人格问卷的效标关联效度分析，例如，研究者想知道君子人格与大五人格有何异同，二者平行，不存在时序问题（葛枭语等，2021）；又如，营销人员常做购物篮分析（关联规则挖掘），了解消费者最

可能在一次购物中同时购买哪个产品和哪个产品（经典案例是啤酒与尿布），无所谓先买哪个、后买哪个，据此可改进货架位置和促销玩法（Hahsler et al., 2005）。（B）预测性研究：常见案例如疾病、心理健康、危险行为的高风险人群筛查预测（例如陈琳等，2024；Hur et al., 2024），招生小组或用人单位利用考试成绩、面试表现、简历资料、人格或心理健康测评去预测入学后或入职后的学业或职业表现等。（C）因果性研究（解释性研究）：占据现在心理学实证研究的主要部分，旨在回答为什么。各目标的区分如图 1。如果目标真的是前二者，没必要从因果关系的视角去考核研究设计的质量；但须指出的是，将本质为因果目标的研究主动降格为预测性研究，但又在全文中明示暗示改变 X 的现实关切（例如建议增减 X 以改善 Y），不应逃脱因果视角的审查（Alvarez-Vargas et al., 2023；Bartram et al., 2024；Rohrer & Wenz, 2024）。不宜认为因果性研究是心理学的唯一目标或高级目标，而对非因果性研究抱以偏见或轻视；契合研究话题的目标才是合适的。

其二，对于因果性研究（解释性研究），即使没有给出明确的因果证据，也可以允许研究者探讨因果意涵。许多文章都有类似的呼吁，例如，Hernán（2018）建议，如果目标是因果性研究，则使用因果语言来准确描述研究目标，不必遮遮掩掩，在结果以外的部分使用因果术语可以是合适的，但在结果部分则忠实地描述结果；Grosz 等人（2020）也建议非实验的研究亦可公开谈论因果效应；类似讨论亦见于 Bailey 等人（2024）、Bartram 等人（2024）。毫无疑问，这并非建议研究者过度夸大因果结论，想要解决的问题是什么、解决到了何种程度，可以通过语言区分开来。一个可能的模板或许是：本文旨在研究收入不平等与幸福感之间的因果关系（不平等影响幸福），在实证上支持了二者的共变性，在理论上论证了二者的方向性，排除了替代解释一二三四（例如排除了经济发展阶段的潜在混淆，确保收入平等与不平等地区的对比并非仅仅反映了贫穷与富裕地区的对比……），但未能排除替代解释五六七八（例如可能存在某种文化价值观氛围既助长不平等、又同时损害幸福，又如幸福感知可能会反过来影响一地居民对不平等的容忍程度……），这些结果使我们对二者的因果关系比较有信心（增进收入平等可以改善公民幸福）或尚不能对因果关系做出评估（仅初步提示了可能性，须待后续研究）。

其三，坦诚沟通非因果性的结果。温忠麟等人（2024）举出的许多带有方向性、但未描述为“影响关系”的语句，似乎并无不妥，例如，association、relationship、linkage、predictable 都是清晰而老实的表达方式，若研究者担心无法表明方向，在 Y 前面加上“后续的”或“几年后的”就可以轻松解决这个担忧；其所举出的另一例句“来自祖先多样性较高地区的人们更愿意公开表达情感”似乎是一个非常不错的表达方式，诚实地描述了所观测到的事实本身。

如果仍有研究者担忧，坦白自己证据的非因果性本质可能招致轻视，不妨看看这个预注册的实验：Alvarez-Vargas 等人（2023）要求 142 名心理学教师、博士后、博士阅读论文摘要，有的人读到的摘要使用了 impact、affect、cause 这样的因果语言，另一些人读到的摘要使用了 predict。结果显示，使用 predict 的摘要被参与者感知到的研究设计和分析质量显著高于使用因果语言的摘要。据此，即便从实用角度出发，正在考虑将自己的非因果性结果切换为“影响关系”的研究者也可能需要三思，因为现有证据似乎未能显露出他们可以从中获益的迹象。

其四，因果证据不是全或无。尤其不宜认为，问卷数据（或更宽泛一些，观察性数据）就注定无法提供因果信息。温忠麟等人（2024）的表述——“心理学界通常认为，基于问卷调查的研究不能研究变量之间的因果关系”——似欠考虑。正如前文举出的一连串方法，实际上，心理学和其它学科研究者正在努力通过多种手段从非操纵性的调查中捕捉可以提供因果意涵的信息（Bartram et al., 2024；Gianicolo et al., 2020；Hammerton & Munafò, 2021；Igelström et al., 2022；Pearl, 2009；Rohrer, 2018, 2024；Wang et al., 2023）。近年来，因果推论的分析技术在许多学科均有发展，心理学研究者可考虑超越心理学的学科范畴、在更广泛的方法论文章的学习中精进因果推论手段（外审专家 1 推荐了相关书籍，感兴趣的读者可参阅官网审稿意见）。这些努力或许无法使一项以因果为目标的研究直接抵达图 1 数轴的 1 端；但是，从 0 到 0.2、从 0.4 到 0.6 的努力同样可以贡献宝贵的见解。例如，Nonkovic 等人（2024）通过兄弟姐妹设计来考察母亲孕期抽烟的影响，发现：不同家庭之间的儿童（基因、成长环境有很大差异），DNA 甲基化的差异只有 6%；相比之下，同一个家庭内的不同儿童（基因和环境相似，但母亲在怀有不同兄弟姐妹期间的抽烟行为恰好不同），DNA 甲基化的差异高达 94%；而且，产前和产后二手烟均无类似效应。在无法随机要求部分孕妇抽烟的现实考虑下，这组观察性数据实际上提供了相当程度的因果信息，尽管仍然可以提出替代解释。因此，不应以一句简单的“相关不等于因果”就忽视随机化实验以外的研究手段对于积累因果知识所做出的贡献。

其五，或许本就不应寄望于单一研究就能为因果理论提供结论性的答案，哪怕是随机化实验。正如 Grosz 等人（2020）一句看似自我循环的陈述——“因果推论是推论性推论（speculative inferences）”——所强调的那样，强求单一研究的直接结果可以立即当作因果结论打包带走，是不合理的信念；任何研究手段都只是为评估因果假设采集证据，而非因果本身。因果效应的三角测量框架指出，三方面的努力应当不断协同和循环：改进假设与理论、从观察性数据中获取估计、通过良好定义的干预措施衡量影响（Bailey et al., 2024）。因此，

无论是在图 1 数轴的偏 0 端提示了一些变量的可能关联，还是在图 1 数轴的偏 1 端尽可能地排除了现有理论所能料想到的所有替代解释，都是朝着因果目标前进的过程，唯有这些努力的积累可以使我们对某个因果命题越来越有信心。

4 结语

温忠麟等人（2024）建设性地将汉语心理学界长期悬而未决的“影响”一词用法问题摆在了聚光灯下，使更多研究者注意到这个问题并加入讨论。此评论希望为这一讨论贡献不同的视角，提醒学界同仁考虑描述相关情形的其它可能办法、审慎决策是否将温忠麟等人（2024）所提出的“影响关系”概念引入研究。

参考文献

Alvarez-Vargas, D., Braithwaite, D., Lortie-Forgues, H., Moore, M., Wan, S., Martin, E., & Bailey, D. H. (2023). Hedges, mottes, and baileys: Causally ambiguous statistical language can increase perceived study quality and policy relevance. *PloS One, 18*(10), e0286403.

Bailey, D. H., Jung, A. J., Beltz, A. M., Eronen, M. I., Gische, C., Hamaker, E. L., ... Murayama, K. (2024). Causal inference on human behaviour. *Nature Human Behaviour, 8*(8), 1448–1459.

Bartram, D., Alaimo, L. S., Avery, E., Bardo, A., Bella, E. D., Binder, M., ... Tani, M. (2024). Towards the next fifty years of Social Indicators Research: Some guidance for authors. *Social Indicators Research, 174*(1), 1–17.

Chen, L., Xu, J., Chen, W., Xu, Y., He, L., & Chai, C. (2024). Differences in HIV-related behaviors among MSM college/university students with different sexual orientations: A cross-sectional survey in Zhejiang province. *Chinese Journal of Public Health, 40*(6), 724–729.

[陈琳, 徐杰, 陈婉君, 徐云, 何林, 柴程良. (2024). 不同性取向大学生男男性行为者艾滋病感染相关行为差异性分析. *中国公共卫生, 40*(6), 724–729.]

Cheng, Y., Feng, Y, Li, Y., Ma, J., Shen, L., Zhang, W, ... Feng, Q. (2023). The developmental trajectory of oral vocabulary knowledge and its predictive effects on reading abilities among Chinese primary school students: A latent growth model. *Acta Psychologica Sinica, 55*(7), 1074–1086.

[程亚华, 冯瑶, 李宜逊, 马嘉琪, 沈岚岚, 张文建, ... 冯秋迪. (2023). 小学儿童口语词汇知识的发展轨迹及其对阅读能力的预测：一个潜变量增长模型. *心理学报, 55*(7), 1074–1086.]

Engstrom, H. R., Laurin, K., Kay, N. R., & Human, L. J. (2024). Socioeconomic status and meta-perceptions: How

markers of culture and rank predict beliefs about how others see us. *Personality and Social Psychology Bulletin, 50*(9), 1386–1407.

Ge, X. (2023). Experimentally manipulating mediating processes: Why and how to examine mediation using statistical moderation analyses. *Journal of Experimental Social Psychology, 109*, 104507.

Ge, X., Li, X., & Hou, Y. (2021). Confucian ideal personality traits (Junzi personality): Exploration of psychological measurement. *Acta Psychologica Sinica, 53*(12), 1321–1334.

[葛枭语, 李小明, 侯玉波. (2021). 孔子思想中的君子人格：心理学测量的探索. *心理学报, 53*(12), 1321–1334.]

Gianicolo, E. A. L., Eichler, M., Muensterer, O., Strauch, K., & Blettner, M. (2020). Methods for evaluating causality in observational studies. *Deutsches Ärzteblatt International, 117*(7), 101–107.

Gladstone, J. J., Matz, S. C., & Lemaire, A. (2019). Can psychological traits be inferred from spending? Evidence from transaction data. *Psychological Science, 30*(7), 1087–1096.

Grosz, M. P., Rohrer, J. M., & Thoemmes, F. (2020). The taboo against explicit causal inference in nonexperimental psychology. *Perspectives on Psychological Science, 15*(5), 1243–1255.

Haber, N. A., Wieten, S. E., Rohrer, J. M., Arah, O. A., Tennant, P. W. G., Stuart, E. A., ... Fox, M. P. (2022). Causal and associational language in observational health research: A systematic evaluation. *American Journal of Epidemiology, 191*(12), 2084–2097.

Hahsler, M., Grün, B., & Hornik, K. (2005). arules - A computational environment for mining association rules and frequent item sets. *Journal of Statistical Software, 14*(15), 1–25.

Hamaker, E. L., Mulder, J. D., & IJzendoorn, M. H. V. (2020). Description, prediction and causation: Methodological challenges of studying child and adolescent development. *Developmental Cognitive Neuroscience, 46*, 100867.

Hammerton, G., & Munafò, M. R. (2021). Causal inference with observational data: The need for triangulation of evidence. *Psychological Medicine, 51*(4), 563–578.

Hernán, M. A. (2018). The C-word: Scientific euphemisms do not improve causal inference from observational data. *American Journal of Public Health, 108*(5), 616–619.

Hur, J. K., Heffner, J., Feng, G. W., Joormann, J., & Rutledge, R. B. (2024). Language sentiment predicts changes in depressive symptoms. *Proceedings of the National Academy of Sciences of the United States of America, 121*(39), e2321321121.

Igelström, E., Craig, P., Lewsey, J., Lynch, J., Pearce, A., & Katikireddi, S. V. (2022). Causal inference and effect

estimation using observational data. *Journal of Epidemiology and Community Health, 76*(11), 960–966.

Kim, S., Lee, H., Lee, J., Lee, S. W., Kwon, R., Kim, M. S., ... Fusar-Poli, P. (2024). Short- and long-term neuropsychiatric outcomes in long COVID in South Korea and Japan. *Nature Human Behaviour, 8*, 1530–1544.

Knight, S., Rocklage, M. D., & Bart, Y. (2024). Narrative reversals and story success. *Science Advances, 19*(34), adl2013.

Liu, Y., Wu, Z., Wang, Y., Dong, Z., Sun, Z., & Gan, Y. (2024). Neoliberalism and governmental and individual responses to the COVID - 19 pandemic: A cross - national analysis. *Political Psychology, 45*(2), 363–382.

Lu, J. G., Jin, P., & English, A. S. (2021). Collectivism predicts mask use during COVID-19. *Proceedings of the National Academy of Sciences of the United States of America, 118*(23), 1–8.

Mann, O., Edin, K. J., & Shaefer, H. L. (2024). Understanding the relationship between intergenerational mobility and community violence. *Proceedings of the National Academy of Sciences of the United States of America, 121*(33), e2309066121.

Martin, M. J., & Donnellan, M. B. (2021). Socioeconomic status, parenting investments, and negative personality traits over time and across generations. *Developmental Psychology, 57*(2), 164–179.

Martin, A., & Kushwaha, T. (2024). Can words speak louder than actions? Using top management teams' language to predict myopic marketing spending. *Journal of Marketing, 88*(6), 140–161.

Meyer, J., Thoma, G.-B., Kampschulte, L., & Köller, O. (2023). Openness to experience and museum visits: Intellectual curiosity, aesthetic sensitivity, and creative imagination predict the frequency of visits to different types of museums. *Journal of Research in Personality, 103*, 104352.

Nonkovic, N., Marceau, K., McGeary, J. E., Ramos, A. M., Palmer, R. H. C., Heath, A. C., & Knopik, V. S. (2024). Maternal smoking during pregnancy is associated with DNA methylation in early adolescence: A sibling comparison design. *Developmental Psychology, 60*(9), 1639–1654.

Pearl, J. (2009). Causal inference in statistics: An overview. *Statistics Surveys, 3*, 96–146.

Pirlott, A. G., & MacKinnon, D. P. (2016). Design approaches to experimental mediation. *Journal of Experimental Social Psychology, 66*, 29–38.

Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science, 1*(1), 27–42.

Rohrer, J. M. (2024). Causal inference for psychologists who think that causal inference is not for them. *Social and Personality Psychology Compass, 18*(3), e12948.

Rohrer, J. M., & Wenz, S. E. (2024). Inappropriate causal assumptions underlie Killingsworth, Kahneman, and Mellers' conclusions. *Proceedings of the National Academy of Sciences of the United States of America, 121*(46), e2313712121.

Rotella, A., Varnum, M. E. W., Sng, O., & Grossmann, I. (2021). Increasing population densities predict decreasing fertility rates over time: A 174-nation investigation. *American Psychologist, 76*(6), 933–946.

Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association, 100*(469), 322–331.

Shen, Y. A., & Shoda, Y. (2021). How candidates' age and gender predict voter preference in a hypothetical election. *Psychological Science, 32*(6), 934–943.

Stasi, M. D., Templeton, E., & Quoidbach, J. (2024). Zooming out on bargaining tables: Exploring which conversation dynamics predict negotiation outcomes. *Journal of Applied Psychology, 109*(7), 1077–1093.

Tam, K. Y. Y., & Inzlicht, M. (2024). Fast-forward to boredom: How switching behavior on digital media makes people more bored. *Journal of Experimental Psychology: General, 153*(10), 2409–2426.

Wang, Q., Wang, Q., & Zhang, R.-Y. (2023). Claim causality with clarity. *Psychoradiology, 3*, kkad007.

Wen, Z., Wang, Y., Ma, P., & Meng, J. (2024). The influence relationship among variables and types of multiple influence factors working together. *Acta Psychologica Sinica, 56*(10), 1462–1470.

[温忠麟, 王一帆, 马鹏, 孟进. (2024). 变量之间的影响关系和多重影响因素的共同作用类型. *心理学报, 56*(10), 1462–1470.]

Wright, A. J., & Jackson, J. J. (2023). Do changes in personality predict life outcomes? *Journal of Personality and Social Psychology, 125*(6), 1495–1518.

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science, 12*(6), 1100–1122.

Yeo, S. C., Lai, C. K. Y., Tan, J., Lim, S., Chandramoghan, Y., Tan, T. K., & Gooley, J. J. (2023). Early morning university classes are associated with impaired sleep and academic performance. *Nature Human Behaviour, 7*(4), 502–514.

The so-called influence relationship requires caution: Commentary on Wen et al. (2024)

GE Xiaoyu

(School of Psychological and Cognitive Sciences and Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing 100871, China)

Abstract

This is a commentary on a paper entitled "The influence relationship among variables and types of multiple influence factors working together" by Wen et al., published in *Acta Psychologica Sinica* in October 2024. They proposed a new concept called the "influence relationship."

This new concept is problematic. First, Wen et al. provided no definition for the "influence relationship," which is unacceptable for a new-conception paper. Second, according to their proposed inference requirement, if researchers fail to disprove alternative explanations that threaten causal inferences, then they can use the term, "influence relationship," when reporting their studies. However, this argument is a manifestation of the misunderstanding of inference requirements of causal relationships. Third, "influence" is a term that poses causal meanings according to Chinese and English dictionaries, previous academic articles, and empirical evidence. Thus, the suggestion by Wen et al. to describe a noncausal relationship using "influence" can result in an overstatement of research significance and misunderstanding among fellow academics and public readers. This scenario is contradictory to the increasing expectations of researchers of more rigorous scientific language. Fourth, Wen et al. were confused with goals and the realization of such goals. Failure to disprove alternative explanations is a compromise or a limitation in methods instead of a unique goal.

Wen et al. stressed that a "directional correlation" lacked an appropriate name in academia. Therefore, they called it the "influence relationship." This

stance is seemingly an unfair description of the academic status quo because researchers typically adopt the word, "predict," to describe a directional correlation. Based on previous articles, this commentary proposes another framework for the categorization of variable relationships. At the goal level, causal goals—in which researchers hypothesize a difference in *Y* if *X* is deliberately changed—can be distinguished from noncausal goals. Furthermore, noncausal goals can be classified as predictive goals (e.g., using texts to predict mental disorder risks or test scores to predict future performance) and purely correlational goals (e.g., a shopping basket analysis or a correlation analysis between a newly proposed personality construct and the Big Five). Neither is concerned with alterations to *X*. At the realization level, if a researcher opts for a causal goal but fails to provide sufficient evidence to support causal relationships, then they are expected to avoid causal language (e.g., "influence") when reporting results and key conclusions. Alternatively, they can use terms such as "be associated with" and "predict" if appropriate.

Moreover, this commentary provides authors and reviewers with several practical suggestions. (A) Clearly define research goals because the different criteria to evaluate causal, predictive, and purely correlational studies should be followed. (B) Enable researchers to discuss causal meanings conveyed by their results even if they fail to offer sufficient causal evidence when targeting causal goals. This statement does not mean an encouragement of overstatement; conversely, only if researchers clearly define their causal goals can they admit the extent to which they are realizing such goals. (C) Use noncausal language to report noncausal results frankly rather than using euphemisms as a strategy for impression management. (D) Avoid an all-or-none attitude toward causal evidence; instead, value every effort that helps disprove alternative explanations and provides more confidence in causal propositions. (E) Do not rely on a single study (even a randomized experiment) to provide conclusive answers to causal questions; instead, value the accumulation of evidence and triangulation.

Keywords influence relationship, correlation relationship, causal relationship, influence factor, predictive research